

て 第1輯』(1924年)などである。このように、関東大震災以外の災害に関する資料や、災害復興に尽力した人物に関連する資料をもデジタルアーカイブの対象に加えると、より便利になると思われる。

他機関の所蔵やデジタルアーカイブとの連携は意識せざるを得ないが、まずは、自館で所蔵する資料の保存と提供にしっかりと責任を持ち、利用者や他機関から信頼されるように努めたい。

- (1) 特集：阪神地域の震災復興・都市問題. 1995, 86(8), p. 2-80.
- (2) 市政専門図書館編. 関東大震災に関連する文献目録. <http://www.timr.or.jp/library/docs/kanto-daishinsai-mokuroku.pdf>. (参照 2013-03-29).
- (3) 倉林義正編. 関東大震災(1923)関連主要文献目録(その1~4). 一橋大学経済研究所. 1982-1985.
- (4) 市政専門図書館デジタルアーカイブ. <http://www.timr.or.jp/library/digitalarchives.html> (参照 2013-03-29).

[受理：2013-05-15]

Yasuhiro Tamura.

Tracing the History of Developing the Collection of the Great Kanto Earthquake at Municipal Reference Library.

CA1793 XXXXXXXXXX 歴史学研究のためのデジタル・アーカイブ —情報発見のために必要なものとは—

ことう まこと
後藤 真*

1. はじめに

本稿では、情報歴史学の立場から、歴史学研究のためのデジタル・アーカイブの現状・検索モデルの課題を通じてデジタル・アーカイブの利用を考えることとする。事例を具体的にするために歴史学研究に関わるものを対象として選ぶが、そこからさまざまな場面でのデジタル・アーカイブの広範な利用へとつながると考えている⁽¹⁾。

2. 研究のための情報検索発見手法とその課題

まずは、歴史学研究者が今まで行ってきた情報を検索し、発見し、成果を発表するまでの手法について一般化してみたい。

1. 研究目的に即した史料「群」を検索する⁽²⁾
2. その史料群を総覧する
3. 総覧した中から重要な情報を抽出する
4. 課題の変化に応じ、新たな史料群を検索する

*花園大学

5. 抽出した史料を再構成、それをもとに論文の叙述を行う

これだけを見ると、歴史学に限らず、多くの研究分野における情報発見とその成果の発表形式であるといえる。しかし、ここで注意しなければならないのは、歴史学研究者の場合、先行研究も史料も必要なものは、「可能である限り、すべてをチェックする」ということである。検索により必要な情報のみを抽出し、それだけをチェックするのではなく、ある種の群単位で史料を検索したのち、その史料群をすべて見ているという点である。これは、歴史学研究の場合、史料によって表記の方法が異なったり、直接的には書かれていないが、その周辺の状況・状態から当該のものを指すと理解できるものなど、既存の検索手法では見つからない場合もあるからである。

しかし、総覧にもおのずから限界が生じるため、その対象をあらかじめ範囲設定するのが一般的である。先行研究であれば、読むべき学術雑誌の範囲があらかじめ決まっており、その範囲を超えるものは、学界のネットワークで情報を得る。史料の場合、とりわけ史料の多い時代であれば、対象とする時代・地域などを経験的に決定したうえで目録を総覧し、その中から史料の中身(コンテンツ)の閲覧へと移行していく。古い時代の場合には、木簡などの新出土資料を除けば、史料の絶対数が限られているため、目録をめくる機会も多くなく、「すべての」コンテンツを見る。ただし、史料の中身だけでは、その性質がはっきりしないものもある⁽³⁾。その場合にはコンテンツを総覧し、そのうえでコンテンツから史料の性質の調査へと進む。この場合も、調査の範囲が経験的に決定されているといえるであろう。

この作業の中では、よほどの偶然を除けば、想定された範囲と全く異なる史料を発見することがわかるだろう。たとえば、近畿のある地域の史料で研究を行おうとするときに、東北や九州など、まったく異なる史料の情報を総覧することは、何らかのヒントが与えられない限りは行わない。

このような既存の情報検索手法の中で、コンピュータでの大量・無差別な検索は、この調査の前提を大幅に変える可能性を秘めたものとして、1990年代末から語られてきた。今までの手法では予想もつかないような情報が発見できるのではないか、との「夢」を持ってきたのである。

しかし現実として、日本史学に限れば、予想外の発見がなされる例は決して多くなかった。史料目録のデータベースは(そもそも数が少ないということを差し引いても)、それ自体が散在していたため、(労力の軽減によって、見る数に変化があったとしても)通常

の目録を総覧する範囲を大きく超えた検索とはなりにくかった。また、たとえ単一の史料の中であってもテキストの全文検索が研究者の総覧の域を超えるに至ることは少なかった。これを解消するために、横断検索の手法が多く検討されたが、ノイズの問題やデータ構造が史料ごとに異なるといった問題もあり、大きな進展となる手法とはなりきれなかった。

3. 予想を超える発見のためのデータの重要性

このような状況の中で、予想もつかないものを発見するためにはどのようなものが求められるだろうか。私は、以下の2点が重要であると考えている。

- ① 検索の対象となるデータ量を増大させると同時に資料構造に基づいたデータモデルを構築すること。
- ② 効果的な知識情報のネットワークを構築させること。

例えば①に関していえば、私は、近代デジタルライブラリーで、正倉院関係資料を検索し活用する事例について発表する機会をいただき、関連した調査をしたことがあった⁽⁴⁾。そのための情報検索の結果、予想していない地域において、正倉院に入っている古文書⁽⁵⁾の写しが掲載されている文献を発見した⁽⁶⁾。そこには、本文情報と関連情報が掲載されており、その情報だけでも正倉院に伝来する古文書の情報が近世から近代にかけてどのように伝播したかを考える重要なきっかけとなり得るものであった。

この事例における正倉院伝来文書の情報は、文献の通常の記事情報の中には存在しないものであり、かつ、正倉院宝物が関連しない地域の博物館の目録に含まれていた。したがって、今までの文献検索の方法では発見できないものであった。近代デジタルライブラリーでは目次の情報が検索対象となっており、かつ、目次で正倉院文書に関連するものが掲載されていると理解できる情報が表示されていることで、はじめて目にするに至ったものである。おそらく、この文書の情報は、近代デジタルライブラリーが目次の情報を検索可能とすると同時に、本文中の内容がわかるような情報の提示手法でなければ、発見できなかったものであろう。このように資料情報を増大させ、構造を記述するだけでも、多くの予想外の発見が可能となる。情報量の増大と整理と効果的な情報提示が、まずは利用者のためになるという一つの例でもあるといえるであろう。

ただし、検索対象とするデータ量は増やせばよいというものではない。事実、データ量の増大は、結果的にノイズを増幅させ、一覧性に乏しい結果をもたらすことになっている。これを回避するためには、コンテンツの情報を効果的に構造化させ、レベルや性質をわ

けて、情報を提示することが欠かせない。デジタル・アーカイブであれば文献の目録情報だけではなく、本文そのものがテキストデータとして検索可能であり、かつ構造化された形で検索可能となることが望まれる。

史料本文をテキストデータとして構造的に記述する手法については、TEI (Text Encoding Initiative) を用いるのが世界標準である⁽⁷⁾。しかし、TEIを用いた日本史学のコンテンツ情報の提供例は、いまだに存在しない。日本でも過去にはTEIへの挑戦はあったのだが、結果的に成功しなかった。しかし、コンテンツの量が増大し、コンテンツの構造化モデルを再考しなければならなくなった昨今、あらためてTEIの重要性を考えるべき時期が来ているといえる。日本史学の史料においてTEIを導入するためには、シンプルなTipsが必要である。日本における史料学・古文学の文脈に即した形でTEIとの関係性をもたせるような事例の構築が必要である。

4. 知識情報ネットワークにもとづく新たな検索発見モデルと Linked Data

もう一つの重要な論点は、知識情報ネットワークモデルの確立である。保立道久氏は東京大学史料編纂所の「歴史知識学の構築」に関連させる形で、「知識ベース・オントロジー」の重要性について述べている⁽⁸⁾。保立氏は、「知識体系を対象的・客体的な知識のみでなく、その客体化としての儀式（あるいは制度）、記録と記憶装置の最奥にある知識生産の現場にまで踏みこんで全体的に復元」⁽⁹⁾するとして、史料上に書かれた文字だけではなく、その背景にあるものを対象として検索可能にすることの重要性を述べたものである。そのうえで、情報学的手法を「便利なもの」としてだけ使うのではなく、学問的方法論として歴史学に内在化する必要があるのではないかと指摘している。

これは先見性の高い議論であり、歴史学研究における史料情報発見を研究に即して端的に述べたものである。

しかし、保立氏の述べたオントロジモデル⁽¹⁰⁾は、安定した技術が実質的に追いつかなかった⁽¹¹⁾。また、モデルの重要性は提起されるものの、実質的にオントロジを「書く」段階にまで進むことができていなかった。その点において、保立氏の議論は情報発見の実践にたどり着くまでには至っていなかった。

近年は、オントロジよりは若干緩やかな形でのネットワークの構築とRDFによって、それを可能とする動きがある。それがLinked Dataの動きである(CA1746参照)。これが、歴史学を含む人文科学にまで適用できる段階に来ており、今後の発見モデルとし

て必要なものになってくるであろう。

歴史学の立場から、Linked Data の最大の利点を端的に述べるならば、「自分でスコープを設定可能であり、同時にスコープを超える可能性を持つ」ということではないだろうか。既存の横断検索は、検索できるデータベースの対象が増えていくというものであった。しかし、Linked Data で構築されるデータ群は、検索できる(狭義の)データベースという概念自体では定義されない。そのため、データベースを複数移動して検索するという発見モデル(≒範囲を絞って目録を総覧するという研究者のモデル)のように研究者の事前の範囲設定とは異なる検索となる点が注目できる。この想定された範囲を超えることと、Linked Data による多様なデータネットワークの確立が、あらたな情報検索発見手法となる可能性を持っている⁽¹²⁾⁽¹³⁾。

5. おわりにかえて

歴史学研究者は、史料から情報を抽出する際、最後には必ず大量の情報を総覧する。その作業自体は、どこまでいっても変わらない。たとえデジタル化が大きく進んでも、その状況が変わることはないであろう。ただ、その総覧する対象や範囲を広げたり「ずらす」可能性をデジタル・アーカイブは秘めている。その点では過去に語られた「夢」を持ち続けている。「予想もつかない史料情報の発見」はコンピュータそのものが行うのではなく、デジタル・アーカイブが、研究者の思考や視角をずらしたその先にあるのではないだろうか。その「ずらし」の可能性をより大きくしていくことが、デジタル・アーカイブを作る側には求められている。それは、保立氏が懸念するような「歴史学が情報学を手段化する」形式⁽¹⁴⁾ではなく、協同の重要な形態であると考えている。学問の内部だけでは研究者の視角を「ずらす」ことは、不可能に近いからである。

- (1) なお、日本史学に関するデジタル・アーカイブに関しては日本歴史学会編『日本歴史』2010年1月号(740号)に「日本史研究とデータベース」という特集が組まれており、2010年時点での状況が確認できる。特集：日本史研究とデータベース。日本歴史。2010。(740)。p. 1-149。
- (2) 本稿での議論を歴史に絞るため、「資料」ではなく「史料」の語を用いる。
- (3) たとえば、正倉院文書は刊行されているテキストデータを見るだけでは、史料の原状がはっきりとしないため、必要な史料について「復原」作業を行う必要がある。
- (4) 後藤真。"利用事例報告「日本古代史研究に近代デジタルライブラリーを使う」"。国立国会図書館。
http://www.ndl.go.jp/jp/dbforum/pdf/slide_kansai2011_3.pdf。(参照 2013-03-31)。
- (5) 厳密には東大寺文書であり、正倉院文書ではない。
- (6) 鎌田共済会。郷土博物館陳列品解説。第4回(昭和4年7月)。鎌田共済会。1929。p. 1-4。
- (7) The Text Encoding Initiative Consortium。"The Text Encoding Initiative"。2013-4-17。<http://www.tei-c.org>。

- (8) 保立道久。歴史のひろば：歴史データベースの将来と歴史知識学：コンピュータはただの便利な道具か。歴史評論。2008。(702)。p. 80-89。
- (9) 前掲。p. 86-87。
- (10) なお、「オントロジー」と「オントロジ」の表記については、3音以上の母音がある場合には最後の長音を省略してもよいというJIS8301規格にもとづき、最後の長音を省略する場合がある。本稿でも引用を除き省略することとする。
- (11) 溝口理一郎著、人工知能学会編『オントロジー工学』(初版、オーム社、2005年)など情報工学の側でも研究は進んでいたが、人文科学の応用へとつながるところまでは困難であった。溝口理一郎。オントロジー工学。人工知能学会編。オーム社。2005。280p。
- (12) 若干蛇足ではあるが、Linked Data の思考の重要な点として、URI の提供もあげられる。決まったポイントが提供されることで、研究者ははじめて Web 上のリソースを典拠として指示できることになる。これが URI の変わらないデータベースでの検索結果である場合、同じリソースを同条件で検索しているという前提が簡単に導かれなため、再現可能性が乏しく、デジタル・アーカイブを研究の典拠に使えないという問題点を抱えていた。Linked Data の発想は、この課題も解決するという点でも必要な方向性である。
- (13) 史料からテキストの文字列に加えて、研究知識情報を抽出し、その情報の意味ネットワークを緩やかに構築することで、あらたな研究発見のモデルを検討するという保立氏の目論見に近いことも、手前味噌ではあるが始めている。正倉院文書にかかわる研究情報をタームとして抽出し、オントロジモデルでデータベースを構築するもので、多様な検索手段を正倉院文書の研究の方法論に即して確保するしくみである。
後藤真ほか。正倉院文書データベースへのトピックマップ応用による奈良時代知識情報構築の試み、じんもんこん 2012 論文集。2012。(7)。p. 125-132。
- (14) 保立。前掲。p. 89。

[受理：2013-05-15]

Makoto Goto.

Digital Archives for Historical Research: What Is Necessary to Discover New Information?.

CA1794

地方議会図書室に明日はあるか —都道府県議会図書室を例に—

つくだ ひろし*
塚田 洋*

1. はじめに

本誌の読者は、地元の議会図書室を訪れたことがあるだろうか。先頃、国立国会図書館(NDL)主催イベントで行われた北川正恭氏(早稲田大学大学院教授、元三重県知事)の講演⁽¹⁾を聴いて、初めて議会図書室に関心を持った方もいるかもしれない。講演の中で北川氏は、知事時代に行った県議会図書室の改革⁽²⁾を例に、図書館関係者のミッションをとらえ直す必要性について指摘した。住民自治への貢献を、広く図書館関係者全体のミッションと位置付ければ、全国の議会図書室の活性化は劇的に進み、そして、このような取組こそが「社会を創る図書館の力」であると述べた。改

*関西館アジア情報課長